Special Source Operations

Content Acquisition Optimization

[redacted]

Yahoo Webmessenger

- Update data sent to individuals logged into Yahoo's Instant Messenger service online
  - Online contact status, unread emails in Yahoo inbox
  - Usually small sessions (2-4kB)
- Sporadic collection (30,000 – 60,000 sessions per day)
- Intermittent bursts of collection against contacts of targets
  - Large numbers of sessions (20,000+) against a single targeted selector
  - Not collected against the target (online presence/unread email from target)
  - No owner attribution (metadata value limited to fact-of comms for emails, online presence events for buddies)
- Over a dozen selectors detasked in two weeks
  - Because a target's contact was using/idling on Yahoo Webmessenger
  - Several very timely selectors (Libyan transition, Greek financial related)

Address Books

- Email address books for most major webmail are collected as stand-alone sessions (no content present*)
- Address books are repetitive, large, and metadata-rich
- Data is stored multiple times (MARINA/MAINWAY, PINWALE, CLOUDs)
- Fewer and fewer address books attributable to users, targets
- Address books account for ~ 22% of SSO's major accesses (up from ~ 12% in August)

Access (10 Jan 12)	Total Sessions	Address Books
US-3171	1488453	237067 (16% of traffic)
DS-200B	938378	311113 (33% of traffic)
US-3261	94132	2477 (3% of traffic)
US-3145	177663	29336 (16% of traffic)
US-3180	269794	40409 (15% of traffic)
US-3180 (16 Dec 11)	289318	91964 (32% of traffic)
TOTAL	3257738	712366 (22% of traffic)

Provider	Collected	Attributed	Attributed%
Yahoo	444743	11009	2.48%
Hotmail	105068	1115	1.06%
Gmail	33697	2350	6.97%
Facebook	82857	79437	95.87%
Other	22881	1175	5.14%
TOTAL	689246	95086	13.80%

Buddy Lists,Inboxes

- Unlike address books, frequently contain content data
  - Offline messages, buddy icon updates, other data included
  - Webmail inboxes increasingly include email content
  - Most collection is due to the presence of a target on a buddy list where the communication is **not** to, from, or about that target
- NSA collects, on a representative day, ~ 500,000 buddylists and inboxes
  - More than 90% collected because tasked selectors identified only as contacts (not communicant, content, or owner)
- Identifying buddylists and inboxes without content (or without useful content) an ongoing challenge

Scenario: [redacted]@yahoo

- [redacted] Sep 2011 [redacted]@yahoo.com (tasked S2E, asw Iran Quds Force) has his/her Yahoo account hacked by an unknown actor, sends out spam email to his/her contact list:

DNI Parser Webmail Display YAHOO! MAIL Active user: [redacted]

Message	
Date	2011-09-[redacted]
Subject	[redacted]!!!!! (New)
From	[redacted]@yahoo.com>
To	[redacted]@yahoogroups.com

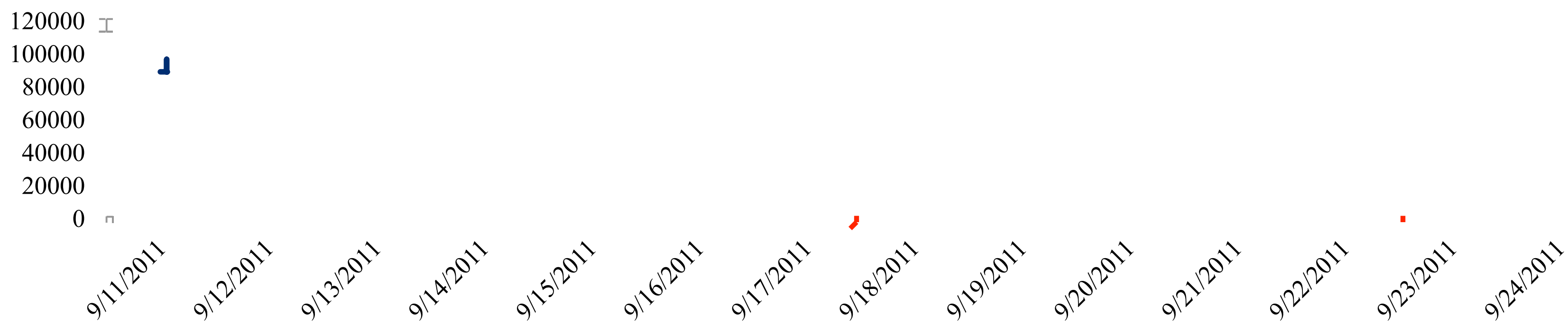
[redacted]

Scenario: [redacted]@yahoo

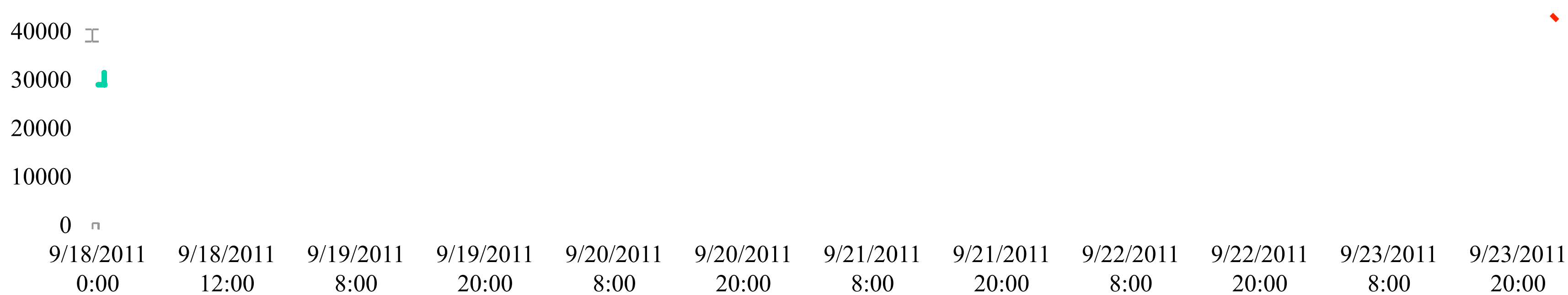
- [redacted]@yahoo.com has a number of Yahoo groups in his/her contact list, some with many hundreds or thousands of members
- At DS-200B in particular, collection spiked as:
  - The initial spam messages were sent (and collected)
  - Inboxes of email recipients were viewed by [redacted] contact list
  - Messages were sometimes viewed, but more often sent as precached views on Google and Yahoo (along with inboxes)
  - Inboxes where the recipient did not delete the spam message continued to be collected every time they were viewed
  - Some recipients added [redacted]@yahoo.com to their address books (possibly as a spam defeat?) – address books were collected every time

Scenario: [redacted]@yahoo

DS-200B Collection By Day - 11 Sep - 24 Sep (in MB)



DS-200B Collection By Hour – 18 Sep – 23 Sep (in MB)



Scenario: [redacted]@yahoo

- [redacted]@yahoo.com emergency detasked from DS-200B and US-3171 at 13:04Z on 20 Oct
- Numerous first-order address books and inboxes collected meant tasked selectors on address books or buddy lists of contacts of [redacted]@yahoo.com also affected:
  - [redacted]@yahoo.com and [redacted]@gmail.com emergency detasked off US-3171 at 13:10Z on 20 Sep
- Memorializing to PINWALE only address books and inboxes owned by target selectors would have reduced PINWALE volumes 90%+
  - Site XKEYSCOREs would buffer data for SIGDEV purposes
  - Metadata from known owner address books and inboxes stored regardless

Mobile IMAP

- IMAP protocol used by email clients to fetch mail from server(s)
- Not designed for devices with intermittent connections (i.e. mobile phones)
- Android implementation in particular uses a lot of bandwidth

```
A0 CAPABILITY
A1 LOGIN [redacted]
A2 CAPABILITY
A3 EXAMINE INBOX
A4 LIST "" INBOX
A5 LIST "" "INBOX.%"
A6 SEARCH SINCE 15-Aug-2011 UNDELETED ALL
A7 FETCH 17 (ENVELOPE INTERNALDATE RFC822.SIZE
A8 FETCH 17 (BODY.PEEK[HEADER])
A9 CLOSE
A10 LOGOUT
```

Date	From	To	Subject	Attachments
Fri Aug [redacted]	[redacted]	[redacted]	2nd Payment Reminder [redacted]	0

Display Information: Email Send to

Subject: 2nd Payment Reminder [redacted]
From: [redacted]
To: [redacted]
Date: Fri Aug [redacted]

Text Size View Full Screen

DNI Parser: Document or message has no data